

Sample of one of the files on submitted disk

10/077, 699A

- (1) GENERAL INFORMATION
- (2) INFORMATION FOR SEQ. ID NO.1:
- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 5001 BASE - #PAIRS
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: SINGLE
- (D) TOPOLOGY: LINEAR
- (ii) MOLECULE TYPE: GENOMIC DNA
- (xi) SEQUENCE DESCRIPTION: SEQ. ID NO.1

Per 1.824 of  
Sequence Rules,  
submit ONE file  
containing all sequences

Does Not Comply  
Corrected Diskette Needed

Sample  
Sequence  
Listing  
Attached

GTTGCTGTTGCTGTTCTAGAACAAATCCATACACACGATTAGATTGAGCTCACCTTCAGCT  
CACGGAAAATTCTTCAGGCCTCAACCCTTCAGCTCCACCCTGCCTTTCTGGAAAAATGCA  
CTCGTGGCTCTACAGGGTGAGCAACCAGGGGCGCAACTGCAGGGCATGCTCATACAGAAC  
ATGCTGCCGAGCTGATCATCGCTCAGCAGTGCAAGCTGCGCACTGGCAGCTTGCA  
TTGTAGCTGGTGTACAACATTCCAGAAGCCGACTGGTATTCGTTGCAATTGTCACAATTG  
TGACGCCCATGCAAGGCCACGAGCAATATCGACTGCAGAACCCTGTGCTGGGATCTACG  
GGAATGATTGGATTGGACGATGTCAGGGCGTTTCGACAGCACCGTACCAAAGCTTGCCAAA  
CTTTAGCAGCGGCTGCTAGCAACCACGAGATAAGCCATGGCCACAACCTTGCAACATCGC  
GCATCTGCAGCCGCCGATGCATGCAAGGTCGGTGTGTGCGGTTTCTGCTTGCTCTGCTT  
CAGGCAACACAGCCTCCAGGTGTTCAACTTGAAGGTGTGACACCACTGGTGTGCTGGCAG  
CTGGCCATTCGGTTTAAAGCCAAGCAGTACAGCGCTGTCAGCTTCATCCCCGCCTGGTTAC  
TGTGATGTATGTGCTTCTGATCAAGCGGTCTCCATGCCGTCCGAACAGAACTGCGCTGT  
AAGCTTACGCAGCCCCAACCGGCTCCGAGCAGCATGCCCTTAAGTGGCGGGAAACTGCC  
AGGGACGGTGTAAGGGCGCCATTACGCGCTCGATACTGTAAGATTGTTTTAGATGAAACA  
GAAATACACCTCCGGAGCTGCGAGTAGCGAGGTGATTTGCATAAGGGATCCACACTGTT  
GTGGGCGCACGTCCAAGAAATGTTTACCCGTTTCGATTGACAGCAAAACATCATGATCAT  
CAAAGGAGTGCATCGACAGTCAACGATCACCAGGTGATTACGTTTGTCACTGACAAGCGC  
CCTCTACGTGCGCCTTGGGCCTACATATGCCCTGCTGTGGGAGTACCCGTGCACAACAGA  
GCGTTAGAGATACTTCATAGCTGCAACTAGACTACCTTTACCCTAACGAAATCACCCTAG  
ACCGACAGTGTGCGAGTAGCTGCGACCCAAACGTGATGGCGAGCGGATTGCTTCTCAAGC  
AGCGCTCGGTATGCCTGAGTGGCAACCGGGAGGTGCGGTATGCTGTTTCTGTCCGCCCGCC  
AGTGAACAGGCGGGCTGTGGTGGCAGCAGGTGCGCTTCTTCTGAAGGGCAGCTAGGGCTG  
TTTCGGGCAGTGCATGCCGGCCTATTTTGGGTTGCTCGGAGCAATAATATGTACTATATT  
GCTCTCGTGGAGCTGTGTTGCGCCACGTGCTTGCTTGGCGCCTGTTGACCCCGGACCCT

This  
is  
invalid  
format.

FYI: all  
U.S.  
applications  
filed on or  
after  
July 1, 1998,  
which  
cannot claim  
a prior  
application  
filed before July 1,  
1998 need to be  
in new Sequence Rules  
format

<110> Smith, John; Smithgene Inc.

<120> Example of a Sequence Listing

<130> 01-00001

<140> PCT/EP98/00001

<141> 1998-12-31

<150> US 08/999,999

<151> 1997-10-15

<160> <

<170> PatentIn version 2.0

<210> 1

<211> 389

<212> DNA

<213> Paramecium sp.

<220>

<221> CDS

<222> (279)...(389)

<300>

<301> Doc. Richard

<302> Isolation and Characterization of a Gene Encoding a  
Protease from Paramecium sp.

<303> Journal of Genes

<304> 1

<305> 4

<306> 1-7

<307> 1988-06-31

<308> 123456

<309> 1988-06-31

<400>	1						
agctgtagtc	attcctgtgt	cctcttctct	ctgggcttct	cacctgtcta	atcagatctc		60
agggagagtg	tcttgacctt	cctctgcctt	tgcagcttca	caggcaggca	ggcaggcagc	E	120
tgatgtggca	attgctggca	gtgccacagg	cttttcagcc	aggcttaggg	tgggttccgc		180
cgcggcgagg	cggccctctt	cgcgctctct	tcgcgcctct	ctctcgtctt	ctcttcgctc		240

## Appendix 3, page 2

ggacctgatt aggtgagcag gaggagggggg cagtttagc atg gtt tca atg ttc agc 296  
Met Val Ser Met Phe Ser

ttg tct ttc aaa tgg cct gga ttt tgt ttg ttt gtt tgt ttg ttc caa 344  
Leu Ser Phe Lys Trp Pro Gly Phe Cys Leu Phe Val Cys Leu Phe Gln

tgt ccc aaa gtc ttc ccc tgt cac tca tca ctg cag ccg aat ctt 389  
Cys Pro Lys Val Leu Pro Cys His Ser Ser Leu Gln Pro Asn Leu

<210> 2  
<211> 37  
<212> PRT  
<213> Paramecium sp.

<<00> 2  
Met Val Ser Met Phe Ser Leu Ser Phe Lys Trp Pro Gly Phe Cys Leu  
1 5 10 15

Phe Val Cys Leu Phe Gln Cys Pro Lys Val Leu Pro Cys His Ser Ser  
20 25 30

Leu Gln Pro Asn Leu  
35

<210> 3  
<211> 11  
<212> PRT  
<213> Artificial Sequence

<220>  
<223> Designed peptide based on size and polarity to act as a linker between the alpha and beta chains of Protein XYZ.

<400> 3  
Met Val Asn Leu Glu Pro Met His Thr Glu Ile  
1 5 10

<210> 4  
<400> 4  
000

[Annex VIII follows]

identifiers and their accompanying information as shown in the following table. The numeric identifier shall be used only in the "Sequence Listing." The order and presentation of the items of information in the "Sequence Listing" shall conform to the arrangement given below. Each item of information shall begin on a new line and shall begin with the numeric identifier enclosed in angle brackets as shown. The submission of those items of information designated with an "M" is mandatory. The submission of those items of information designated with an "O" is optional. Numeric identifiers <110> through <170> shall only be set forth at the beginning of the "Sequence Listing." The following table illustrates the numeric identifiers.

Numeric Identifier	Definition	Comments and Format	Mandatory (M) or Optional (O)
<110>	Applicant	Preferably max. of 10 names; one name per line; preferable format: Surname, Other Names and/or Initials	M
<120>	Title of Invention		M
<130>	File Reference	Personal file reference	M, when filed prior to assignment of appl. number
<140>	Current Application Number	Specify as: US 07/999,999 or PCT/US96/99999	M, if available
<141>	Current Filing Date	Specify as: yyyy-mm-dd	M, if available
<150>	Prior Application Number	Specify as: US 07/999,999 or PCT/US96/99999	M, if applicable include priority documents under 35 USC 119 and 120
<151>	Prior Application Filing Date	Specify as: yyyy-mm-dd	M, if applicable
<160>	Number of SEQ ID NOs	Count includes total number of SEQ ID NOs	M
<170>	Software	Name of software used to create the Sequence Listing	O
<210>	SEQ ID NO: #:	Response shall be an integer representing the SEQ ID NO shown	M
<211>	Length	Respond with an integer expressing the number of bases or amino acid residues	M

<212>

Type

Whether presented sequence molecule is DNA, RNA, or PRT (protein). If a nucleotide sequence contains both DNA and RNA fragments, the type shall be "DNA." In addition, the combined DNA/RNA molecule shall be further described in the <220> to <223> feature section.

M

<213>

Organism

Scientific name, i.e. Genus/species, Unknown or Artificial Sequence. In addition, the "Unknown" or "Artificial Sequence" organisms shall be further described in the <220> to <223> feature section.

M

<220>

Feature

Leave blank after <220>. <221-223> provide for a description of points of biological significance in the sequence.

M, under the following conditions: if "n," "Xaa," or a modified or unusual L-amino acid or modified base was used in a sequence; if ORGANISM is "Artificial Sequence" or "Unknown"; if molecule is combined DNA/RNA.

<221>

Name/Key

Provide appropriate identifier for feature, preferably from WIPO Standard ST.25 (1998), Appendix 2, Tables 5 and 6

M, under the following conditions: if "n," "Xaa," or a modified or unusual L-amino acid or modified base was used in a sequence

<222>

Location

Specify location within sequence; where appropriate state number of first and last bases/amino acids

M, under the following conditions: if "n," "Xaa," or a modified or unusual L-amino acid or modified

		in feature	base was used in a sequence
<223>	Other Information	Other relevant information; four lines maximum	if, under the following conditions: if "n," "Xaa," or a modified or unusual L-amino acid or modified base was used in a sequence; if ORGANISM is "Artificial Sequence" or "Unknown"; if molecule is combined DNA/RNA.
<300>	Publication Information	Leave blank after <300>	0
<301>	Authors	Preferably max. of ten named authors of publication; specify one name per line; preferable format: Surname, Other Names and/or Initials	0
<302>	Title		0
<303>	Journal		0
<304>	Volume		0
<305>	Issue		0
<306>	Pages		0
<307>	Date	Journal date on which data published; specify as yyyy-mm-dd, MM-yyyy or Season-yyyy	0
<308>	Database Accession Number	Accession number assigned by database including database name	0
<309>	Database Entry Date	Date of entry in database; specify as yyyy-mm-dd or MM-yyyy	0
<310>	Patent Document Number	Document number; for patent-type citations only. Specify as, for example, US 07/999,999	0

<311>	Patent Filing Date	Document filing date, for patent-type citations only; specify as yyyy-mm-dd	0
<312>	Publication Date	Document publication date, for patent-type citations only; specify as yyyy-mm-dd.	0
<313>	Relevant Residues	FROM (position) TO (position)	0
<400>	Sequence	SEQ ID NO should follow the numeric identifier and should appear on the line preceding the actual sequence	M

5. Section 1.024 is revised to read as follows:

1.024 Form and format for nucleotide and/or amino acid sequence submissions in computer readable form.

(a) The computer readable form required by 1.021(c) shall meet the following specifications:

(1) The computer readable form shall contain a single "Sequence Listing" as either a diskette, series of diskettes, or other permissible media outlined in paragraph (c) of this section.

(2) The "Sequence Listing" in paragraph (a) (1) of this section shall be submitted in American Standard Code for Information Interchange (ASCII) text. No other formats shall be allowed.

(3) The computer readable form may be created by any means, such as word processors, nucleotide/amino acid sequence editors or other custom computer programs; however, it shall conform to all specifications detailed in this section.

(4) File compression is acceptable when using diskette media, so long as the compressed file is in a self-extracting format that will decompress on one of the systems described in paragraph (b) of this section.

(5) Page numbering shall not appear within the computer readable form version of the "Sequence Listing" file.

(6) All computer readable forms shall have a label permanently affixed thereto on which has been hand-printed or typed: the name of the applicant, the title of the invention, the date on which the data were recorded on the computer readable form, the operating system used, a reference number, and an application serial number and filing date, if known.

(b) Computer readable form submissions must meet these format requirements:

(1) Computer: IBM PC/XT/AT, or compatibles, or Apple Macintosh;

(2) Operating System: MS-DOS, Unix or Macintosh;

10/077699

- (1) GENERAL INFORMATION
- (2) INFORMATION FOR SEQ. ID NO.1:
  - (i) SEQUENCE CHARACTERISTICS:
    - (A) LENGTH: 5001 BASE - #PAIRS
    - (B) TYPE: NUCLEIC ACID
    - (C) STRANDEDNESS: SINGLE
    - (D) TOPOLOGY: LINEAR
  - (ii) MOLECULE TYPE: GENOMIC DNA
  - (xi) SEQUENCE DESCRIPTION: SEQ. ID NO.1

GTTGCTGTTGCTGTTCTAGAACAAATCCATACACACGATTAGATTGAGCTCACCTTCAGCT  
CACGGAAAAATTCTTCAGGCCTCAACCCTTCAGCTCCACCCTGCCTTTCTGGAAAAATGCA

(sample of one file on submitted disk)

The above is in invalid format for a U.S. application  
Sequence Listing. Per 1.824 of Sequence Rules, "The  
computer readable form shall contain a single 'Sequence  
Listing' as either a diskette..." DO not create  
multiple files for a Sequence Listing. Also, see  
sample Sequence Listing (attached) for valid  
format. Suggestion: Please consult Sequence Rules.



<110> Smith, John; Smithgene Inc.  
 <120> Example of a Sequence Listing  
 <130> 01-00001  
 <140> PCT/EP98/00001  
 <141> 1998-12-31  
 <150> US 08/999,999  
 <151> 1997-10-15  
 <160>   
 <170> Patent in version 2.0  
 <210> 1  
 <211> 389  
 <212> DNA  
 <213> Paramecium sp.  
 <220>   
 <221> CDS  
 <222> (279)...(389)  
 <300>   
 <301> Doe, Richard  
 <302> Isolation and Characterization of a Gene Encoding a  
 Protease from Paramecium sp.  
 <303> Journal of Genes  
 <304> 1  
 <305> 4  
 <306> 1-7  
 <307> 1988-06-31  
 <308> 123456  
 <309> 1988-06-31  
 <400> 1  
 agctgtagtc attcctgtgt cctcttctct ctgggcttct caccctgtca atcagatctc 60  
 agggagagtg tcttgacctt cctctgcctt tgcagcttca caggcaggca ggcaggcagc E 120  
 tgaatgtggca attgctggca gtgccacagg ctttccagcc aggccttaggg tgggttccgc 180  
 cgcggcgagg cggccctctt cgcgctcttc tcgcgctctt ctctcgctct cctctcgctc 240

Please consult

Appendix 3, page 2

ggacctgatt aggtgagcag gaggagggggg cagtttagc atg gtt tca atg ttc agc 296  
Met Val Ser Met Phe Ser

ttg tct ttc aaa tgg cct gga ttt tgt ttg ttt gtt tgt ttg ttc caa 344  
Leu Ser Phe Lys Trp Pro Gly Phe Cys Leu Phe Val Cys Leu Phe Gln

tgt ccc aaa gtc ttc ccc tgt cac tca tca ctg cag ccg aat ctt 389  
Cys Pro Lys Val Leu Pro Cys His Ser Ser Leu Gln Pro Asn Leu

<210> 2  
<211> 37  
<212> PRT  
<213> Paramecium sp.

<<00> 2  
Met Val Ser Met Phe Ser Leu Ser Phe Lys Trp Pro Gly Phe Cys Leu  
1 5 10 15

Phe Val Cys Leu Phe Gln Cys Pro Lys Val Leu Pro Cys His Ser Ser  
20 25 30

Leu Gln Pro Asn Leu  
35

<210> 3  
<211> 11  
<212> PRT  
<213> Artificial Sequence

<220>  
<221> Designed peptide based on size and polarity to act as a linker between the alpha and beta chains of Protein XYZ.

<400> 3  
Met Val Asn Leu Glu Pro Met His Thr Glu Ile  
1 5 10

<210> 4  
<400> 4  
000

[Annex VIII follows]

identifiers and their accompanying information as shown in the following table. The numeric identifier shall be used only in the "Sequence Listing." The order and presentation of the items of information in the "Sequence Listing" shall conform to the arrangement given below. Each item of information shall begin on a new line and shall begin with the numeric identifier enclosed in angle brackets as shown. The submission of those items of information designated with an "M" is mandatory. The submission of those items of information designated with an "O" is optional. Numeric identifiers <110> through <170> shall only be set forth at the beginning of the "Sequence Listing." The following table illustrates the numeric identifiers.

Numeric Identifier	Definition	Comments and Format	Mandatory (M) or Optional (O)
<110>	Applicant	Preferably max. of 10 names; one name per line; preferable format: Surname, Other Names and/or Initials	M
<120>	Title of Invention		M
<130>	File Reference	Personal file reference	M, when filed prior to assignment of appl. number
<140>	Current Application Number	Specify as: US 07/999,999 or PCT/US96/99999	M, if available
<141>	Current Filing Date	Specify as: yyyy-mm-dd	M, if available
<150>	Prior Application Number	Specify as: US 07/999,999 or PCT/US96/99999	M, if applicable include priority documents under 35 USC 119 and 120
<151>	Prior Application Filing Date	Specify as: yyyy-mm-dd	M, if applicable
<160>	Number of SEQ ID NOs	Count includes total number of SEQ ID NOs	M
<170>	Software	Name of software used to create the Sequence Listing	O
<210>	SEQ ID NO: #:	Response shall be an integer representing the SEQ ID NO shown	M
<211>	Length	Respond with an integer expressing the number of bases or amino acid residues	M

<212>	Type	Whether presented sequence molecule is DNA, RNA, or PRT (protein). If a nucleotide sequence contains both DNA and RNA fragments, the type shall be "DNA." In addition, the combined DNA/RNA molecule shall be further described in the <220> to <223> feature section.	M
<213>	Organism	Scientific name, i.e. Genus/species, Unknown or Artificial Sequence. In addition, the "Unknown" or "Artificial Sequence" organisms shall be further described in the <220> to <223> feature section.	M
<220>	Feature	Leave blank after <220>. <221-223> provide for a description of points of biological significance in the sequence.	M, under the following conditions: if "n," "Xaa," or a modified or unusual L-amino acid or modified base was used in a sequence; if ORGANISM is "Artificial Sequence" or "Unknown"; if molecule is combined DNA/RNA.
<221>	Name/Key	Provide appropriate identifier for feature, preferably from WIPO Standard ST.25 (1998), Appendix 2, Tables 5 and 6	M, under the following conditions: if "n," "Xaa," or a modified or unusual L-amino acid or modified base was used in a sequence
<222>	Location	Specify location within sequence; where appropriate state number of first and last bases/amino acids	M, under the following conditions: if "n," "Xaa," or a modified or unusual L-amino acid or modified

		in feature	se was used in sequence
<223>	Other Information	Other relevant information; four lines maximum	M, under the following conditions: if "n," "Xaa," or a modified or unusual L-amino acid or modified base was used in a sequence; if ORGANISM is "Artificial Sequence" or "Unknown"; if molecule is combined DNA/RNA
<300>	Publication Information	Leave blank after <300>	0
<301>	Authors	Preferably max of ten named authors of publication; specify one name per line; preferable format: Surname, Other Names and/or Initials	0
<302>	Title		0
<303>	Journal		0
<304>	Volume		0
<305>	Issue		0
<306>	Pages		0
<307>	Date	Journal date on which data published; specify as yyyy-mm-dd, MM-yyyy or Season-yyyy	0
<308>	Database Accession Number	Accession number assigned by database including database name	0
<309>	Database Entry Date	Date of entry in database; specify as yyyy-mm-dd or MM-yyyy	0
<310>	Patent Document Number	Document number; for patent-type citations only. Specify as, for example, US 07/999,999	0

<311>	Patent Filing Date	Document filing date, for patent-type citations only; specify as yyyy-mm-dd	
<312>	Publication Date	Document publication date, for patent-type citations only; specify as yyyy-mm-dd	0
<313>	Relevant Residues	FROM (position) TO (position)	0
<400>	Sequence	SEQ ID NO should follow the numeric identifier and should appear on the line preceding the actual sequence	M

5. Section 1.024 is revised to read as follows:

1.024 Form and format for nucleotide and/or amino acid sequence submissions in computer readable form.

(a) The computer readable form required by 1.021(c) shall meet the following specifications:

(1) The computer readable form shall contain a single "Sequence Listing" as either a diskette, series of diskettes, or other permissible media outlined in paragraph (c) of this section.

(2) The "Sequence Listing" in paragraph (a) (1) of this section shall be submitted in American Standard Code for Information Interchange (ASCII) text. No other formats shall be allowed.

(3) The computer readable form may be created by any means, such as word processors, nucleotide/amino acid sequence editors or other custom computer programs; however, it shall conform to all specifications detailed in this section.

(4) File compression is acceptable when using diskette media, so long as the compressed file is in a self-extracting format that will decompress on one of the systems described in paragraph (b) of this section.

(5) Page numbering shall not appear within the computer readable form version of the "Sequence Listing" file.

(6) All computer readable forms shall have a label permanently affixed thereto on which has been hand-printed or typed: the name of the applicant, the title of the invention, the date on which the data were recorded on the computer readable form, the operating system used, a reference number, and an application serial number and filing date, if known.

(b) Computer readable form submissions must meet these format requirements:

(1) Computer: IBM PC/XT/AT, or compatibles, or Apple Macintosh;

(2) Operating System: MS-DOS, Unix or Macintosh;